

The Honorable Lauren King

UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF WASHINGTON

PUMA SE, PUMA NORTH AMERICA, INC.,
and LLOYD IP LIMITED,

Plaintiffs,

v.

BROOKS SPORTS, INC.,

Defendant.

Case No. 2:23-cv-00116-LK

**PLAINTIFFS PUMA SE AND PUMA
NORTH AMERICA, INC.'S MOTION
TO EXCLUDE TESTIMONY AND
OPINIONS OF SARAH BUTLER**

NOTED FOR CONSIDERATION ON:
November 13, 2024

ORAL ARGUMENT REQUESTED

PLAINTIFFS' MOTION TO EXCLUDE SARAH BUTLER

QUARLES & BRADY LLP
Attorneys at Law
411 East Wisconsin Avenue
Suite 2400
Milwaukee, Wisconsin 53202-4428
414-277-5000

TABLE OF CONTENTS

		Page
1		
2		
3	SUMMARY OF ARGUMENT	1
4	KEY FACTS	1
5	ARGUMENT	3
6	I. Surveys Must be Relevant and Conducted According to Accepted Principles.....	3
7	II. The Butler Survey Is Irrelevant to Forward Confusion	4
8	a. Forward Confusion Surveys Show Senior User Stimuli First	4
9	b. The Butler Survey Does Not Follow The Recognized Procedure for Testing Forward Confusion.....	6
10	III. The Butler Survey is Inadmissible Because it Was Not Conducted Using Accepted Principles	7
11	a. To Be Admissible, a Survey Must Be Conducted Using Acceptable Survey Principles	7
12	b. Reliable Surveys Replicate Marketplace Conditions.....	8
13	c. Butler’s Survey Did Not Replicate Marketplace Conditions By Using An Altered Test Stimulus	9
14	d. Reliable Surveys Do No Manipulate Results By Changing Questions For Some Participants and Not Others.....	10
15	e. Butler Improperly Changed the Survey Questions	12
16	f. Reliable Surveys Use Controls Unrelated to the Tested Term	12
17	g. Butler’s Survey Used an Improper Control	13
18	CONCLUSION.....	15
19		
20		
21		
22		
23		
24		
25		
26		

TABLE OF AUTHORITIES

Cases

<i>American Footwear Corp. v. General Footwear Co. Ltd.</i> , 609 F.2d 655 (2d Cir. 1979).....	9
<i>Bach v. Forever Living Prod. U.S., Inc.</i> , 473 F. Supp. 2d 1110 (W.D. Wash. 2007).....	3
<i>CytoSport, Inc. v. Vital Pharm., Inc.</i> , 894 F. Supp. 2d 1285 (E.D. Cal. 2012).....	3, 6
<i>Daubert v. Merrell Dow Pharm., Inc.</i> , 509 U.S. 579 (1993).....	3, 15
<i>Government Employees Ins. Co. v. Google, Inc.</i> , No. 1:04CV507, 2005 WL 1903128 (E.D. Va. Aug. 8, 2005)	13
<i>Gucci Am., Inc. v. Guess?, Inc.</i> , 831 F. Supp. 2d 723 (S.D.N.Y. 2011).....	9
<i>Herman Miller, Inc. v. Belnick LLC</i> , No. 1:18-cv-5012-WMR, 2021 WL 2582547 (N.D. Ga. Jan. 11, 2021)	5
<i>Icon Enters. Int’l, Inc. v. Am. Prods. Co.</i> , No. CV 04-1240 SVW (PLAx), 2004 WL 5644805 (C.D. Cal. Oct. 7, 2004).....	4, 7, 15
<i>J & J Snack Foods, Corp. v. Earthgrains Co.</i> , 220 F. Supp. 2d 358 (D.N.J. 2002)	4
<i>Jackpocket, Inc. v. Lottomatrix NY LLC</i> , 645 F. Supp. 3d 185 (S.D.N.Y. 2022), <i>aff’d</i> , No. 23-12-CV, 2024 WL 1152520 (2d Cir. Mar. 18, 2024)	5
<i>Kargo Global, Inc. v. Advance Magazine Publishers, Inc.</i> , No. 06 CIV 550 JFK, 2007 WL 2258688 (S.D.N.Y. Aug. 6, 2007)	8, 9, 12, 15
<i>Kournikova v. Gen. Media Comm’cns, Inc.</i> , 278 F. Supp. 2d 1111 (C.D. Cal. 2003)	10
<i>Kudos, Inc. v. Kudoboard, LLC</i> , No. 20-cv-1876-SI, 2021 WL 5415258 (N.D. Cal. Nov. 20, 2021)	5
<i>Luckenbach Texas, Inc. v. Engel</i> , No. A-19-cv-00567-DH, 2022 WL 16857410 (W.D. Tex. Oct. 13, 2022)	5
<i>M2 Software, Inc. v. Madacy Entm’t</i> , 421 F.3d 1073 (9th Cir. 2005)	3, 7

1	<i>Mountaineers Found. v. Mountaineers,</i>	
	No. 2:19-cv-1819-RSL-TLF, 2022 WL 1061925 (W.D. Wash. Apr. 8, 2022)	3
2	<i>Oracle America, Inc. v. Google, Inc.,</i>	
3	No. C 10-03561 WHA, 2016 WL 1743116 (N.D. Cal. May 2, 2016).....	11
4	<i>Reinsdorf v. Skechers U.S.A.,</i>	
	922 F. Supp. 2d 866 (C.D. Cal. 2013)	7
5	<i>Scott Fetzer Co. v. House of Vacuums Inc.,</i>	
6	381 F.3d 477 (5th Cir. 2004)	15
7	<i>Squirtco v. Seven-Up Co.,</i>	
	628 F.2d 1086 (8th Cir. 1980)	2
8	<i>Survivor Media, Inc. v. Survivor Prods.,</i>	
9	406 F.3d 625 (9th Cir. 2005)	4
10	<i>THOIP v. Walt Disney Co.,</i>	
	690 F. Supp. 2d 218 (S.D.N.Y. 2010).....	8, 15
11	<i>Trevino v. Gates,</i>	
12	99 F.3d 911 (9th Cir. 1996)	3
13	<i>Trouble v. Wet Seal, Inc.,</i>	
	179 F. Supp. 2d 291 (S.D.N.Y. 2001).....	8
14	<i>U.S. Polo Ass’n, Inc. v. PRL USA Holdings, Inc.,</i>	
15	800 F. Supp. 2d 515 (S.D.N.Y. 2011),	
	<i>aff’d</i> , 511 F. App’x 81 (2d Cir. 2013).....	12, 13
16	<i>Wendt v. Host Int’l, Inc.,</i>	
17	125 F.3d 806 (9th Cir. 1997)	3

Other Authorities

18	5 McCarthy on Trademarks and Unfair Competition § 32:173 (5th Ed.).....	2
19	5 McCarthy on Trademarks and Unfair Competition § 32:174 (5th Ed.).....	2
20	5 McCarthy on Trademarks and Unfair Competition § 32:177 (5th ed.)	4
21	5 McCarthy on Trademarks and Unfair Competition § 32:187 (5th Ed.).....	12
22		
23		
24		
25		
26		

SUMMARY OF ARGUMENT

Brooks' survey expert, Sarah Butler, should be excluded because the methodology she used does not follow accepted principles of survey design and implementation and thus does not meet the requirements of Federal Rules of Evidence 702 and 403. Specifically, the survey contains four critical flaws. First, because of the order of the stimuli used in the survey, the survey does not measure forward confusion, which is the type of confusion at issue. While the order of the stimuli alone is a basis to exclude the survey as irrelevant to the legal issues at hand, this single issue is compounded by three other structural flaws that combine to make the survey and Butler's related opinions about the survey wholly inadmissible.

As a second critical error, Butler changed the survey questions partway through fielding the survey and improperly merged the data from respondents who saw different versions of the survey. By improperly pooling the resulting data, the final measure is unreliable and biased in favor of Brooks.

As a third flaw, Butler doctored the Brooks website used as a test stimulus, such that it did not reflect actual marketplace conditions.

As a fourth flaw, Butler's survey uses a control that improperly contained the disputed term, further skewing the data once the results from this improper control were used in Butler's calculation of the likelihood of consumer confusion.

When each of these four are viewed together, the cumulative impact is a survey that is irrelevant to the forward confusion claim at issue and unreliable. Any possible probative value is outweighed by the substantial danger of unfair prejudice, confusion of the issues, and the potential to mislead the jury. The Butler Survey and opinions are inadmissible and should be excluded.

KEY FACTS

Butler's survey and opinions purport to measure a likelihood of confusion relevant to PUMA's affirmative trademark infringement claim. PUMA has brought claims of trademark infringement and unfair competition under the Lanham Act and common law associated with its

1 use of NITRO on footwear. (Dkt. 187 at ¶¶ 56–64; 70–123.) PUMA advances a theory of forward
 2 confusion: consumers encountering Brooks’ unauthorized use of the NITRO mark are likely to be
 3 confused and believe that PUMA is the source of Brooks’ products. (*See id.* at ¶¶ 59, 60, 73, 83,
 4 84, 93, 94, 105, 118, and 119.) PUMA does not allege reverse confusion, in which consumers
 5 encountering PUMA’s products would mistakenly believe that Brooks is the source.

6 Butler “surveyed 400 consumers who purchase performance running shoes costing \$100
 7 or more.” (Expert Report of Sarah Butler (“Butler Report”), at 4, attached as Ex. 1 to the
 8 Declaration of Lauren Bolcar in Support of Motion to Exclude Expert Report (“Bolcar Decl.”).)
 9 Butler designed a *Squirt*¹ format survey. (Deposition of Sarah Butler (“Butler Dep.”) 238:10–13,
 10 Bolcar Decl. Ex. 2.) Respondents were shown one of two versions of a Brooks advertisement: the
 11 test stimulus which used the NITRO mark, or the control stimulus, which replaced the NITRO
 12 mark with the word “nitrogen.” (Butler Report, at 16–18.) All respondents were then shown three
 13 running shoe product pages: PUMA, HOKA, and Nike. (*Id.* at 18–19.) None of the participants
 14 were shown PUMA’s mark before being shown the Brooks advertisement.

15 Butler had to revise certain questions in her survey after she reviewed the results of the
 16 first 106 respondents and realized based on the responses given to the open-ended questions in the
 17 survey that at least 22 respondents were comparing the wrong images. (Butler Report, fn. 28;
 18 Butler Dep. at 38:12–23; 43:8–25.) Butler removed 22 respondents (over twenty percent) from this
 19 initial set and kept the responses of the remaining 84 respondents in the final data set. (Butler Dep.
 20 at 44:1–5.)

21
 22
 23
 24 ¹ There are two main survey formats for testing likelihood of consumer confusion, *Squirt* and *Eveready*. 5 McCarthy
 25 on Trademarks and Unfair Competition (“McCarthy”) § 32:173 (5th Ed.). The *Squirt* survey format shows survey
 26 respondents both the trademark owner’s and alleged infringer’s marks and products. *See Squirtco v. Seven-Up Co.*,
 628 F.2d 1086 (8th Cir. 1980). The *Eveready* survey shows respondents the allegedly infringing product and asks
 non-leading questions about the company affiliated with the mark, without being shown the senior mark.
 5 McCarthy § 32:174.

ARGUMENT

I. Surveys Must be Relevant and Conducted According to Accepted Principles.

Under the Supreme Court’s *Daubert* decision, a trial court is required to act as a “gatekeeper,” to “ensur[e] that an expert’s testimony both rests on a reliable foundation and is relevant to the task at hand.” *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 597 (1993). The trial court has broad discretion “to admit or exclude expert testimony; expert evidence is properly excluded where ‘foundational facts demonstrating relevancy . . . are not sufficiently established [.]’” *Mountaineers Found. v. Mountaineers*, No. 2:19-cv-1819-RSL-TLF, 2022 WL 1061925, at *2 (W.D. Wash. Apr. 8, 2022) (quoting *Trevino v. Gates*, 99 F.3d 911, 922 (9th Cir. 1996) (internal citations omitted)). Under Ninth Circuit law, a likelihood of confusion survey is only admissible when there is “a proper foundation for admissibility” and when the survey is “relevant and conducted according to accepted survey principles.” *M2 Software, Inc. v. Madacy Entm’t*, 421 F.3d 1073, 1087 (9th Cir. 2005). “A district court must exclude evidence that is either not ‘relevant,’ or not conducted according to accepted principles.” *CytoSport, Inc. v. Vital Pharm., Inc.*, 894 F. Supp. 2d 1285, 1291 (E.D. Cal. 2012) (citing *Wendt v. Host Int’l, Inc.*, 125 F.3d 806, 814 (9th Cir. 1997)). As the party offering the survey, Brooks has the burden to show it is admissible. *M2 Software*, 421 F.3d at 1089.

Pursuant to Federal Rule of Evidence 702, “expert testimony is admissible if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.” *Bach v. Forever Living Prod. U.S., Inc.*, 473 F. Supp. 2d 1110, 1114 (W.D. Wash. 2007). Evidence that is otherwise admissible may be excluded under Federal Rule of Evidence 403 “if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury.” *Id.* at 1114–15. Where, as here, the survey is not aligned with the issues of the case, it is not only irrelevant, but inadmissible under Rule 403 because of the likelihood that the survey findings will confuse the jury. *See, e.g., J & J Snack*

1 *Foods, Corp. v. Earthgrains Co.*, 220 F. Supp. 2d 358, 370 (D.N.J. 2002) (excluding consideration
 2 of a survey at the summary judgment stage that was inconsistent with the legal claims at issue);
 3 *Icon Enters. Int’l, Inc. v. Am. Prods. Co.*, No. CV 04-1240 SVW (PLAx), 2004 WL 5644805, at
 4 *28 (C.D. Cal. Oct. 7, 2004) (citing *Earthgrains* and excluding a flawed survey because “admitting
 5 the survey would raise questions that should not properly be before the jury”).

6 **II. The Butler Survey Is Irrelevant to Forward Confusion**

7 **a. Forward Confusion Surveys Show Senior User Stimuli First.**

8 “Forward confusion occurs when consumers believe that goods bearing the junior mark
 9 came from, or were sponsored by, the senior mark holder.” *Survivor Media, Inc. v. Survivor*
 10 *Prods.*, 406 F.3d 625, 630 (9th Cir. 2005). It is well-established that in a Squirt or similar two
 11 room/line-up survey format measuring the likelihood of forward confusion, the senior user’s
 12 mark is shown first, followed by the junior user’s mark. As explained by Prof. Jacob Jacoby,
 13 “[w]hen both plaintiff’s and defendant’s marks are presented to the same respondent, to parallel
 14 the temporal sequence generally occurring in the marketplace, the sounder rationale for tests of
 15 forward confusion would appear to be to present the first comer’s goods first and the second
 16 comer’s goods second.” Jacoby, Jacob, *TRADEMARK SURVEYS VOL. 1: DESIGNING,*
 17 *IMPLEMENTING, AND EVALUATING SURVEYS* (2013) at 545 (“Jacoby”, Bolcar Decl., Ex. 3.)

18 Likewise, the oft-cited treatise, McCarthy on Trademarks and Unfair Competition,
 19 explains that in a line-up survey method to test forward confusion, “respondents are shown the
 20 plaintiff’s trade dress, and then, after a short delay, shown a line-up of other brands, including
 21 the accused product. Respondents are asked if any of them are made by the same company as
 22 makes the product initially seen.” 5 McCarthy on Trademarks and Unfair Competition
 23 (“McCarthy”) § 32:177 (5th Ed.); *see also* Swann, Jerre B., “Eveready and Squirt—Cognitively
 24 Updated,” 106 *TRADEMARK REP.* 727, 739–40 (2016) (explaining that in a two-room/line-up
 25 study, “in one room, a respondent sees the allegedly infringed brand . . . then sees in a second
 26

1 room a line-up of brands typically in the same product category, that includes the allegedly
 2 infringing brand”). (Bolcar Decl., Ex. 4.)

3 Courts throughout the country have likewise recognized and accepted the method of
 4 showing survey participants the senior user’s mark followed by the junior user’s mark in surveys
 5 seeking to gain information about forward confusion. Points below summarize four district court
 6 cases using nearly the same language to describe the admissible order of stimuli used in surveys
 7 seeking to address cases involving forward confusion:

8 • *Jackpocket, Inc. v. Lottomatrix NY LLC*, 645 F. Supp. 3d 185, 224–
 9 25 (S.D.N.Y. 2022), *aff’d*, No. 23-12-CV, 2024 WL 1152520 (2d Cir. Mar. 18,
 10 2024) (“In a *Squirt* survey, respondents are first exposed to the senior mark and
 11 then to the allegedly infringing mark.”);

12 • *Luckenbach Texas, Inc. v. Engel*, No. A-19-cv-00567-DH, 2022 WL
 13 16857410, at *2 (W.D. Tex. Oct. 13, 2022) (“When conducting a survey using the
 14 *Squirt* format, consumers are first shown the plaintiff’s product or mark, and later
 15 separately shown an array of other products or marks that includes the accused
 16 infringing one and then asked if they think anything from the array is associated
 17 with the first image that they saw.”);

18 • *Herman Miller, Inc. v. Belnick LLC*, No. 1:18-cv-5012-WMR, 2021
 19 WL 2582547, at *1 (N.D. Ga. Jan. 11, 2021) (same);

20 • *Kudos, Inc. v. Kudoboard, LLC*, No. 20-cv-1876-SI, 2021 WL
 21 5415258, at *11 (N.D. Cal. Nov. 20, 2021) (describing an expert’s *Squirt* survey,
 22 during which “respondents were shown screenshot of [the plaintiff’s] website
 23 homepage, which prominently displays the [plaintiff’s] mark. . . . Next, respondents
 24 were shown a randomized series of webpage screenshots from four other companies
 25 that offer employee recognition and engagement software products, including [the
 26 defendant’s] business-level product webpage.”).

Butler has previously applied the standard format of showing the senior user's mark first for a test of forward confusion. In October 2023, Butler submitted an expert report in a different case involving allegations of forward confusion in which she described the "standard and accepted designs for measuring likelihood of confusion" as "respondents are first shown the senior user's mark and subsequently asked to identify other company(ies) that are associated with the senior user." (Bolcar Decl., Ex. 5 at 18.) Butler even cited to Prof. Jacoby to support her position: "As Jacoby explains, an appropriately designed two-room Squirt survey 'expose[s] respondents to the first comer's stimulus first...followed by exposing respondents to the allegedly infringing stimulus second [].'" (*Id.* at fn. 61 (quoting Jacoby, at 563).)

b. The Butler Survey Does Not Follow The Recognized Procedure for Testing Forward Confusion.

PUMA's only theory of confusion is forward confusion, i.e., that consumers are likely to be confused when encountering Brooks' (the junior user) unauthorized use of NITRO and believe that PUMA (the senior user) provided, approved, or sponsored Brooks' products. Butler agrees that PUMA has alleged forward confusion. (Butler Dep. 185:20–22.)

Butler designed a *Squirt* format survey but failed to show PUMA's mark first. (Butler Dep. 238:10–13.) Butler's Survey contravenes case law, respected treatises, and Butler's own prior writings and surveys regarding the proper format for a likelihood of confusion survey testing forward confusion. The Survey thus departed from accepted survey design principles, rendering it and the accompanying report unreliable, wholly irrelevant to PUMA's claims that exclusively allege forward confusion, and inadmissible. *See e.g. CytoSport, Inc.*, 894 F. Supp. 2d, at 1291 (excluding the defendant's expert survey because the defendant "cite[d] no authority that shows that the methods used by [the expert] are generally acceptable.")

When questioned in her deposition about the order of the stimuli her survey used, Butler attempted to diminish the importance of the proper order of survey stimuli but in doing so mischaracterized her own survey. Butler asserts that her survey asked "the forward confusion

questions” of “do you think that Brooks’ brand or the company is associated, connected to, or sponsored by Puma.” (Butler Dep. at 183:1–13.) Not true. The survey questions presented to respondents were far more ambiguous. For example, Q3 asked whether respondents believed the advertisement (Brooks test or control stimulus) they were shown first and the running shoes from the product page (of PUMA or a distractor) were made or put out by the same brand or company or different brands or companies. (Butler Report, Ex. D at 9.) Accordingly, respondents could have believed that Brooks made or put out the PUMA shoes (reverse confusion) or that PUMA made or put out the Brooks shoes (forward confusion).

Because the Butler Survey does not test forward confusion, which is the legal theory PUMA alleged in its complaint, the survey results are not relevant to the claims in this case and are inadmissible under Rule 702. The survey is also properly excluded under Rule 403, as admitting the survey “would raise questions that should not properly be before the jury.” *Icon Enters. Int’l, Inc.*, 2004 WL 5644805, at *28 (excluding a survey whose “probative value is substantially outweighed by its potential to confuse and mislead the jury.”).

III. The Butler Survey is Inadmissible Because it Was Not Conducted Using Accepted Principles.

a. To Be Admissible, a Survey Must Be Conducted Using Acceptable Survey Principles.

“Unless survey evidence is conducted according to accepted principles, it is not admissible in the first instance.” *Reinsdorf v. Skechers U.S.A.*, 922 F. Supp. 2d 866, 878 (C.D. Cal. 2013). The proponent of a survey must establish that it was conducted according to accepted principles for the survey to be admissible under Rule 702. *M2 Software*, 421 F.3d at 1089. When a proffering party does “not identify any scientific principles underlying the [expert’s] survey” and the survey on its face “appears to violate numerous accepted practices in the field of survey research” the survey should be excluded. *See Reinsdorf*, 922 F. Supp. at 878 (granting motion to exclude survey expert testimony and report).

1 Recognized survey principles include using stimuli that replicate marketplace conditions,
 2 reporting data and performing calculations in a reliable manner, and using a control that effectively
 3 measures survey noise unrelated to the trademark in dispute.

4 **b. Reliable Surveys Replicate Marketplace Conditions.**

5 To provide an accurate indication of likely consumer confusion, a survey must reflect how
 6 consumers encounter the relevant goods and services in the marketplace. “Typically, trademark
 7 infringement surveys use stimuli, such as pictures, advertisements or clothing, that directly expose
 8 potential consumers to the products or the marks in question.” *Trouble v. Wet Seal, Inc.*, 179 F.
 9 Supp. 2d 291, 308 (S.D.N.Y. 2001). “The failure of a survey to approximate actual marketplace
 10 conditions can provide grounds for inadmissibility.” *THOIP v. Walt Disney Co.*, 690 F. Supp. 2d
 11 218, 231 (S.D.N.Y. 2010).

12 Courts exclude surveys as unreliable when the test stimulus is flawed. For example, in
 13 *Kargo Global, Inc. v. Advance Magazine Publishers, Inc.*, the court excluded the plaintiff’s
 14 survey expert’s likelihood of confusion survey and accompanying testimony because the survey
 15 was “so flawed that its probative value is substantially outweighed by its potential for unfair
 16 prejudice and the likelihood that it will confuse or mislead the jury.” No. 06 CIV 550 JFK, 2007
 17 WL 2258688, at *7 (S.D.N.Y. Aug. 6, 2007). The court found that “the survey (1) employed a
 18 format that failed to approximate real world conditions and was impermissibly leading, and (2)
 19 used improper stimuli.” *Id.* As the court explained, “[t]o be probative of actual confusion, a
 20 survey must use stimuli that approximate what a potential customer would encounter in making
 21 purchasing decisions.” *Id.* Thus, a survey using “stimuli that differ from what a consumer is
 22 actually likely to see in the marketplace does not accurately test for actual consumer confusion”
 23 and “lacks probative value.” *Id.* (internal quotation marks and citation omitted). The survey in
 24 *Kargo* used an advertisement that “would never be encountered by [the plaintiff’s] prospective
 25 end users, who were the respondents of the . . . survey.” *Id.* at 11. In addition, the stimuli of the
 26 defendant’s product were found to be “not sufficiently representative.” *Id.* Finally, the court

1 noted that the testimony of the plaintiff's survey expert would prove to be "both powerful and
2 misleading" to the jury and determined that "[t]he application of Rule 403 is clearly warranted in
3 this case." *Id.* at 12.

4 **c. Butler's Survey Did Not Replicate Marketplace Conditions By Using An**
5 **Altered Test Stimulus.**

6 Butler used a flawed test stimulus that failed to accurately depict Brooks' use of the NITRO
7 mark. Butler did not even attempt to recreate realistic marketplace conditions but instead used a
8 doctored Brooks webpage with multiple uses of the disputed mark removed. This unacceptable
9 test stimulus renders the Butler Survey unreliable and inadmissible. *See Gucci Am., Inc. v. Guess?,*
10 *Inc.*, 831 F. Supp. 2d 723, 739 (S.D.N.Y. 2011) ("The failure of a survey to approximate actual
11 marketplace conditions can provide grounds for inadmissibility."); *American Footwear Corp. v.*
12 *General Footwear Co. Ltd.*, 609 F.2d 655, 660 n. 4 (2d Cir. 1979) (holding district court decision
13 to exclude survey "for failure to conduct it under actual marketing conditions" not clearly
14 erroneous).

15 Butler's Survey used a truncated and edited version of Brooks' website as the test stimulus.
16 The actual Brooks website referenced in PUMA's complaint contains two additional uses of
17 NITRO that are not present in the test stimulus. (*Compare* Butler Report Ex. F, pg. 2 *with* Butler
18 Report Ex. G, pg. 4.)

19 Further compounding the issue, the test stimulus used for respondents who completed the
20 survey on a mobile phone included the phrase "Our innovative nitro-infusion process" which does
21 not appear in the test stimulus for respondents who completed the survey on a desktop, laptop, or
22 tablet computer. (*Compare* Butler Report. Ex. E at 82 *with* Butler Report. Ex. F at 2.) Butler
23 testified that this phrase was "inadvertently" left in this version. (Butler Dep. at 88:12–17.) As a
24 result, respondents who completed the survey on a mobile phone and saw the test stimulus were
25 exposed to an additional instance of "nitro" in the webpage content than respondents who viewed
26 the test stimulus on a desktop, laptop, or tablet computer.

Butler also removed references to “nitrogen” that appeared on the webpage. (Butler Report at fn. 23.) At her deposition, Butler repeatedly contended that her survey was “not a test of Brooks’ website.” (Butler Dep. at 68:16–23; 76:4–13; 77:2–78:2.) Yet, Butler later described her survey as designed to “test Puma’s allegation that people could simply be exposed to Brooks’ use of nitro and then somehow be confused and make an association [].” (*Id.* at 84:25–85:5.) And Butler agreed that consumers who visit the real Brooks webpage would see these additional uses of NITRO that were removed from the test stimulus. (*Id.* at 83:5–13.) Butler’s test stimulus is therefore improper because it failed to approximate actual marketplace conditions and artificially reduced respondents’ exposure to Brooks’ use of the NITRO mark. *See Kournikova v. Gen. Media Comm’ns, Inc.*, 278 F. Supp. 2d 1111, 1125 (C.D. Cal. 2003) (excluding a consumer perception survey in part because it “did not permit those sampled to view all of the relevant evidence bearing on the relevant issue.”).

**d. Reliable Surveys Do No Manipulate Results By Changing Questions For
Some Participants and Not Others.**

Survey “questionnaires are data-gathering instruments. And just like any other measuring instrument—be it a thermometer or a scale for measuring our weight—reliance upon the data gathered by a questionnaire is questionable if the data are not scientifically reliable (that is, consistent, repeatable) and valid (yield correct and true answers).” Jacoby at 596. “[A] fundamental requirement of measuring instruments is *standardization*. This means requiring that the instrument and, with the exception of questions that are skipped for some respondents but not others, especially the questions, need to be the same for everyone being measured.” *Id.* at 597 (emphasis in original).

As explained by PUMA’s rebuttal expert, Dr. Justin Anderson, “[w]hen an initial questionnaire is changed to clarify questions known to cause confusion among a significant portion of respondents, it is good practice to start over with a revised version and to remove all of the initial respondents from the survey and subsequent analysis.” (Expert Report of Dr. Justin

1 Anderson (“Anderson Report”), at 18, attached as Ex. A to the Declaration of Dr. Justin Anderson.)
2 This is because “data collected in the pretest . . . could be inaccurate or biased compared to the
3 results of the full-scale study.” Erin Ruel et al., *THE PRACTICE OF SURVEY RESEARCH: THEORY*
4 *AND APPLICATIONS* 117 (2016) (Bolcar Decl., Ex. 6.) While “it may be unreasonable to exclude
5 [pretest] participants from the entire study,” if those respondents are to be included in the final
6 data set, “comparison and discussion of the differences between the pretested groups and the full-
7 scale group is necessary. It is also important to exercise caution when interpreting these results,
8 and it is important to note this potential data contamination as a possible limitation of the research.”
9 *Id.*

10 Courts have recognized these teachings. In *Oracle America, Inc. v. Google, Inc.*, the court
11 considered whether including in the final results of a survey respondents who saw an initial version
12 and respondents who saw a modified version of the survey rendered the results unreliable. No. C
13 10-03561 WHA, 2016 WL 1743116 at *10 (N.D. Cal. May 2, 2016). There, as here, the expert
14 admitted that he modified one of the survey questions “because the pretest suggested that it was
15 *misinterpreted* by some respondents, and another entirely new question was added without
16 explanation.” *Id.* (emphasis in original). The court explained that whether “the pretest results
17 agreed with the overall results of the survey is beside the point. The issue is not whether the pretest
18 results accorded with the full-scale survey results, but whether both were achieved using uniform
19 methodology so as to produce *reliably* similar results.” *Id.* (emphasis in original). The court also
20 criticized the expert’s inclusion of the “pretest results in his overall results without any comparison
21 or discussion of differences between the pretest and full-scale groups, or acknowledgment of how
22 this inclusion might have limited the survey’s reliability or conclusions.” *Id.* The court struck any
23 portion of the expert’s survey or opinions based on pretest results and instructed that the expert
24 “may still refer to the survey’s size, statistical significance, or respondents, but in doing so he must
25 refer only to the full-scale survey, and he must modify any specific numerical findings
26 accordingly.” *Id.*

e. Butler Improperly Changed the Survey Questions.

Contrary to generally accepted survey principles, Butler included in her final results the responses obtained from the initial version of her survey, thus mixing data from survey participants who saw different surveys. Butler admits that she modified three of the survey questions because, after reviewing the results of the first 106 respondents, she determined that some respondents had “misinterpreted” those questions. (Butler Report, fn. 28; Butler Dep. at 38:12–23; 43:8–25.) She removed 22 respondents (over twenty percent) from this initial set and kept the responses of the remaining 84 respondents in the final data set. (Butler Dep. at 44:1–5.)

This violated established survey practice. Moreover, Butler did not evaluate or qualify the results of this change to determine how it may have impacted the data set. However, Dr. Anderson *did* compare the difference in net likelihood of confusion measured in the respondents who saw the original survey and whose data was retained in the final report with the net likelihood of confusion of respondents who saw the revised report. (Anderson Report at 19.) Initial respondents had a net likelihood of confusion of -3.5% while revised respondents had a net likelihood of confusion of 7.9%, a difference of 11.4%. (*Id.* at 19–20.) Dr. Anderson determined that this “constitutes a substantial difference in the survey results between initial and revised respondents.” (*Id.* at 20.) By including the responses of the initial respondents in the final analysis, the level of confusion measured by the Butler Report was biased in favor of Brooks. (*Id.*) This is yet another flaw in the Butler Survey. These flaws significantly outweigh the Survey’s probative value, and the Butler Survey should be excluded. *Kargo*, 2007 WL 2258688, at *7.

f. Reliable Surveys Use Controls Unrelated to the Tested Term.

To be reliable, a survey must include a control to test for “noise” or bias in the design of the survey itself. 5 McCarthy § 32:187. “Without a proper control, there is no benchmark for determining whether a likelihood of confusion estimate is significant or merely reflects flaws in the survey methodology.” *U.S. Polo Ass’n, Inc. v. PRL USA Holdings, Inc.*, 800 F. Supp. 2d 515, 534 (S.D.N.Y. 2011), *aff’d*, 511 F. App’x 81 (2d Cir. 2013).

1 “In designing a survey-experiment, the expert should select a stimulus for the control group
 2 that shares as many characteristics with the experimental stimulus as possible, *with the key*
 3 *exception of the characteristic whose influence is being assessed.*” Diamond, Shari S., “Reference
 4 Guide on Survey Research” *Reference Manual on Scientific Evidence*, 3rd ed., (2011), p. 399
 5 (emphasis added) (Bolcar Decl., Ex. 7.) Courts have criticized controls that were too similar to the
 6 issue being tested. For example, in *U.S. Polo Association*, following a bench trial, the court gave
 7 no weight to a survey which used improper controls that “included the very elements being
 8 addressed, namely the word mark ‘POLO’ and . . . also a mounted polo player image.” 800 F.
 9 Supp. 2d at 534; *see also Government Employees Ins. Co. v. Google, Inc.*, No. 1:04CV507, 2005
 10 WL 1903128, at *5 (E.D. Va. Aug. 8, 2005) (criticizing control for not removing the allegedly
 11 infringing elements that were tested in the main survey. The control “did not function as an
 12 accurate measure of the confusion caused by the non-infringing elements”).

13 **g. Butler’s Survey Used an Improper Control.**

14 Yet another flaw in the Butler Survey is the control stimulus. Butler’s Survey failed to use
 15 a control according to accepted survey principles. To create the control stimulus, Butler started
 16 with the Brooks webpage and “replaced ‘nitro’ with the word ‘nitrogen.’” (Butler Report, at 18.)
 17 Butler agrees that “nitro” is subsumed within “nitrogen.” (Butler Dep, at 101:12–13 (“So if you’re
 18 simply asking N-I-T-R-O is part of the word ‘nitrogen,’ yes.”).) Thus, Butler’s control
 19 unacceptably included the very mark she was assessing. Because “nitrogen” includes the disputed
 20 mark NITRO, Butler’s edit is contrary to the well-established principles of survey design.

21 Butler’s position that “nitrogen” is an appropriate control because it is descriptive of the
 22 nitrogen-infused foam in Brooks’ shoes (*id.* at 101:14–102:9) is contradicted by her own prior
 23 work. In 2007, Butler co-authored an article titled “Emerging Issues in the Use of Surveys in
 24 Trademark Infringement on the Web” in which she wrote “[i]f the survey requires a control, the
 25 researcher must ensure that the control stimulus is properly chosen to be similar in nature to the
 26 relevant trademark, but *without any of the infringing elements at issue in the case.*” (Bolcar Decl.

1 Ex. 8., at 6 (emphasis added).) In other likelihood of confusion surveys designed by Butler, the
2 term in dispute was either removed entirely from the control image (*id.*, Ex. 9) or swapped with a
3 completely different term. (*Id.*, Ex. 10 (Butler survey testing whether services provided by
4 RockStar Hotels, Inc. were associated with Hard Rock Hotels or its brand, in which “RockStar”
5 was changed to “Celebrity” in survey control stimulus.))

6 Butler’s erroneous inclusion of the disputed mark in the control was not harmless. It
7 affected the survey results. Butler admits that, of the respondents who viewed the Brooks control
8 stimulus and thought the Brooks control was put out, made by, or affiliated or associated with
9 PUMA’s shoes, twelve specifically referenced “nitro” or the names of the products. *See, e.g.*,
10 Butler Dep. at 217:11–17 (“So this person made some sort of connection between the Nitro of the
11 Puma ad and the nitrogen of the Brooks ad, correct? A. This respondent made an association
12 between the Brooks control ad shown that said ‘nitrogen,’ and the Puma shoes shown that shown
13 that’s called ‘Nitro.’); 219:8–11 (“This respondent wrote ‘Because they both said nitro on the
14 advertisement.’ Do you see that? A. Yes.”); 220:17–21 (“Q. And then the row above it, 377,
15 respondent 5477. That respondent wrote ‘Both have Nitro.’ Do you see that? A. Yes.”); 225:22–
16 25 (“Q. And this respondent wrote, in the open-ended response, ‘They both have nitro in the
17 running shoe.’ A. That’s correct.”). Again, these were respondents who viewed the control image,
18 which did not contain the word “nitro” but rather “nitrogen.” Clearly these respondents made a
19 connection between the “nitrogen” of the control Brooks advertisement with PUMA’s NITRO
20 mark. Butler found that only 5.5 percent of respondents “associated Brooks and Puma.” (Butler
21 Report at 6.) She obtained this value by subtracting the percent of confusion measured in the
22 control group from the percent of confusion measured in the test group. (*Id.* at 29.) However, the
23 control measurement was inflated due to Butler’s unreliable control stimulus, leading to a lower
24 overall measurement of confusion than might otherwise have been calculated. (Anderson Report,
25 at 31–32.) At bottom, Butler’s control was not designed according to accepted survey principles
26 and provides no reliable means for assessing noise or bias. As a result, “it is not a reliable indicator

1 of consumer confusion” and should be excluded. *THOIP*, 690 F. Supp. 2d 241 (finding a survey
 2 of likelihood of consumer confusion inadmissible because it “failed to replicate actual marketplace
 3 conditions in which consumers encountered the products at issue . . . and failed to use an adequate
 4 control [.]”).

5 While each of these deficiencies, standing alone, may go to the weight of the survey and
 6 not its admissibility, all of these flaws taken together place the survey outside the realm of accepted
 7 principles and render it inadmissible. “[S]erious flaws in a survey will make any reliance on that
 8 survey unreasonable.” *Icon Enters. Int’l Inc.*, 2004 WL 5644805, at *22 (quoting *Scott Fetzer Co.*
 9 *v. House of Vacuums Inc.*, 381 F.3d 477, 488 (5th Cir. 2004)). Additionally, where, as here, the
 10 case will be before a jury, “live testimony of [a] highly seasoned and impressively credentialed
 11 consumer confusion expert, regarding the results of [a] deeply flawed [survey], would prove to be
 12 ‘both powerful and quite misleading.’” *Kargo*, 2007 WL 2258688, at *12 (quoting *Daubert*, 509
 13 U.S. at 595).

14 CONCLUSION

15 The methodology Butler employed flouted generally accepted principles of survey design,
 16 rendering her survey so fundamentally flawed and her conclusions so unreliable that they cannot
 17 fairly be presented to the jury as expert testimony. The Butler Survey used an improper survey
 18 format to test PUMA’s allegation of forward confusion and is thus not relevant to the claims at
 19 issue. In violation of accepted survey principles, the Butler Survey included the disputed mark in
 20 the control stimulus and doctored the test stimulus in a way that did not approximate marketplace
 21 conditions. It included data obtained from respondents who saw a different initial version of the
 22 Survey, resulting in a final measure of likelihood of confusion biased in favor of Brooks. To
 23 prevent the introduction of unreliable and irrelevant evidence and to avoid confusion of the jury
 24 and issues at trial, the Court should exclude Butler’s Report and testimony pursuant to Federal
 25 Rules of Evidence 403 and 702.

I certify that this memorandum contains 5,305 words, in compliance with the Local Civil Rules.

Dated: October 16, 2024

QUARLES & BRADY LLP

By: /s/ Mark P. Walters
ATTORNEY

Mark P. Walters
LOWE GRAHAM JONES PLLC
1325 Fourth Avenue, Suite 1130
Seattle, WA 98101
walters@lowegrahamjones.com
Tel: 206-381-3300

Johanna M. Wilbert (admitted pro hac vice)
Michael T. Piery (admitted pro hac vice)
QUARLES & BRADY LLP
411 East Wisconsin Ave., Suite 1400
Milwaukee, WI 53202
johanna.wilbert@quarles.com
michael.piery@quarles.com
Tel: 414-277-5000

Lauren C. Bolcar (admitted pro hac vice)
QUARLES & BRADY LLP
2020 K Street NW, Suite 400
Washington, DC 20006
lauren.bolcar@quarles.com
Tel: 202-372-9600

Attorneys for PUMA SE and PUMA North America, Inc.